Title
The Case for Enzymatic Competitive Metal Affinity Methods.

Journal
ACS catalysis, 10(3)

ISSN
2155-5435

Authors
Reilley, David J
Hennefarth, Matthew R
Alexandrova, Anastassia N

Publication Date
2020-02-01

DOI
10.1021/acscatal.9b04831

Peer reviewed

The Case for Enzymatic Competitive Metal Affinity Methods

David J. Reilley,[†] Matthew R. Hennefarth,[†] and Anastassia N. Alexandrova[*,†,‡]

[†]*Department of Chemistry and Biochemistry, University of California, Los Angeles, 607 Charles E. Young Drive East, Los Angeles, CA 90095-1569, USA*

[‡]*California NanoSystems Institute, University of California, Los Angeles, 570 Westwood Plaza, Los Angeles, California 90095-1569, USA*

E-mail: ana@chem.ucla.edu

Phone: +1 310 8253769

We often want to know which metal will bind to a protein most readily, which metal or metals actually bind *in vivo*, and which one will be the best at enzymatic catalysis. It is not guaranteed that a single metal could satisfy all of the above for a given natural metalloenzyme. For artificial metalloenzymes (ArMs), we also want to know if the protein can bind the desired metal, and if the metal would then function as a catalyst with the desired activity and selectivity. Hence, being able to compute the metal binding affinities to proteins is desirable in the studies of enzymatic catalysis and enzyme design. Unfortunately, this goal is non-trivial. Efforts toward solving this problem are the focus of this article.

The first step to determine metal affinity is to identify the metal binding site, yet as this is already firmly established for the functional metals of many interesting systems we will not extensively cover it here. Other papers discuss the development of computational tools to address this particular problem for unstudied, poorly resolved, or less accessible biomolecules. While some of these methods can start from a sequence,[1,2] all eventually require some sort

of structure to identify possible binding sites.[3–5] With this constraint, X-ray crystallography remains the most reliable and broadly applicable approach for proteins, even if costly. As with protein folds in general, crystal structures gathered over the last 50 years provide the most likely binding site for a broad range of proteins. As many natural metalloenzymes bind their strongly held metals alongside specifically tailored cofactors, substrates, and scaffolds the likelihood of other significant binding sites is frequently minimal. However, static structures determined for a predominant metal do not answer all questions of metal affinity and function.

The questions of which metal is used in a natural enzyme, and which metal we want to employ in an artificial enzyme are not easily answerable because different forces drive the evolution of enzymes in nature than the priorities of man-made catalysts. Instead of maximizing enzyme activity, biology caps it to maintain the complex equilibria of homeostasis. Biology prioritizes the bio-availability of the starting materials and fold stability, but also ensures that enzymes can be readily destroyed when needed. These constraints also apply to the way in which metals are selected for natural metalloenzymes.[6,7] Furthermore, the catalytically relevant metals for many metalloproteins are not truly known. Many enzymes are assumed to be Zn(II)-dependent based on X-ray crystal structures, but this can be an artifact of experimental conditions.[8] Follow up studies on systems such as histone deacetylase[9,10] carbonic anhydrase,[8] S-ribosylhomocysteinase,[11] and peptide deformylase[12] show that sometimes other metals can bind and report significant activity. In some cases, the metal reported by crystallography is not even a particularly significant contributor to the proteins function. Without considering the binding affinity of different metals, *in vitro* and computational studies of metalloproteins could be based on a false or incomplete picture of metal preferences.

A major goal in the design of artificial metalloenzymes is maximal catalytic performance, with less emphasis on stability in their simpler *in vitro* environment of operation. Previous efforts already found that while proteins provide powerful platforms for new catalysts, the

reactions they can perform, and sometimes their catalytic rates, have hard limitations.[13,14] Recently, directed evolution has become an indispensable tool to develop new ArMs or refine existing ones.[15–19] However, directed evolution is constrained by the roles for which a given protein scaffold has evolved.[13,14,20] While there is promiscuity of function in many proteins, some reactions are simply out of reach of conventional methodologies. Metals that are not natively bioavailable can expand the space of accessible reactions. For example, recent efforts show that noble metals can expand the repertoire of porphyrin-dependent enzymes.[21] However, non-physiological metals must bind sufficiently strongly to their protein scaffolds, whose amino acids did not originally evolve to ligate non-physiological metals. Thus, determination of metal affinity is required. Additionally, as we will show shortly, the affinity of the metal to the protein (e.g. the stability gain upon metal binding) and the catalytic activity may follow a non-trivial and non-linear mutual dependence, via the Brønsted-Evans-Polanyi (BEP) relation.

Lastly, metal-protein affinity is of broader interest than biocatalysis. It is relevant to metal transport about the body, particularly the activity of metal chaperones, which, unlike many proteins, bind metals in a highly selective manner and in specific environments.[22–25] Chaperones help maintain the distinct metal concentrations in different organ systems, tissues, and even different subcellular organelles within cells.[26] Tracking the metal affinity of these proteins in different contexts is important for metal toxicology. A large number of transition and heavy metals are now bioavailable with their use in modern industries, including industrial catalysis. Some metals, such as Cd(II), Hg(II), As(III), and Pb(II) are highly toxic and lead to non-specific syndromes.[27,28] The extent of cytotoxicity of other metals, such as Al(III), Ti(IV), and Ga(III), is unclear but demands investigation as they are introduced into the body both from the environment, and for medical purposes.[29–33] Metal binding may even play a role in neurodegenerative diseases, hypothetically facilitating the protein-protein aggregation and fibril formation.[34] Ultimately it is of high interest to know the metal-protein affinity, and have ways to calculate it.

1 Existing Methods and Their Limitations

Dedicated computational tools to investigate protein-metal binding, which we will refer to as competitive metal affinity (CMA) methods, are hard to come by. The ideal CMA would incorporate an accurate energy evaluation and significant dynamical sampling to capture configurational entropy in order to fully describe the thermodynamics of metal binding. Clearly, the expense of the accurate energy calculations severely limits the amount of sampling that can be afforded. While there are many methods to study metalloprotein behavior in general, not all are suited to form the basis of a CMA method.

Classical force field based methods can be parameterized to model some metalloenzyme structures, but are typically insufficient to obtain thermodynamic values. Force field parameters for metals are based on a point charge supplemented with various harmonic terms and operate on the basis of a fixed metal coordination (e.g. octahedral, tetrahedral) that cannot change significantly as a function of protein dynamics. These potentials can contain bonding and non-bonding interactions, but are generally fitted to capture structure (within limits) rather than energy.[35–38] In this respect they can be fairly successful for systems containing closed shell metals with ideal geometries (Zn(II), Mg(II), Mn(II), Cd(II)), remaining stable over long molecular dynamics (MD) simulations and providing some thermodynamic data.[39–41] However, even the most successful applications of these methods do not obtain reliable energies for catalytic studies.

Electronic structure calculations are necessary to obtain accurate metal binding energies. One possible approach is to use a small cluster model of the active site and treat it quantum mechanically. However, this approach ignores the entropy of the protein scaffold and the impact of the protein dynamics on the energy and entropy of the active site. The only portion of the entropy in the free energy of the active site that this approach captures is the vibrational entropy — typically calculated within the harmonic approximation and subject to the constraints imposed by the rest of the protein structure. While cluster models are useful for catalytic mechanism mapping,[42,43] and as such can play a role in artificial

metalloenzyme design,[44–46] these applications rely on the cancellation of errors when protein entropy is ignored equivalently throughout the reaction profile. On the other hand, many metal exchange phenomena are inaccessible to the approach, as enzymes frequently undergo some amount of restructuring when a new metal binds.

A more promising avenue to obtain metal binding free energies based on electronic structure calculations are mixed QM/MM simulations. This class of methods combines a quantum mechanical description of the metal center and its surrounding environment and molecular mechanical modeling of the rest of the protein (Figure 1). Statistical mechanical sampling of the protein becomes possible within QM/MM, and there has been intensive research into and development of these methods over the last two decades.[47–49] Sufficient sampling is still a problem, however, for most established QM/MM methods. Our group developed the QM/DMD method,[50] which combines DFT with discrete molecular dynamics (DMD)[51] for enhanced sampling. DMD is based on simplified square-well potentials, ballistic equations of motion, and slight coarse graining, which permit extensive conformational sampling of the full protein. QM/DMD divides the protein into three regions: a QM only region comprising the metal center(s) and ligating atoms, a DMD only region comprising the bulk of the protein, and a region treated with both QM and DMD made up of the rest of the active site (normally composed of 50 to 200 atoms). The overlapping region modeled with both theories allows the geometric and energetic information to pass between QM and DMD calculations. Typically, one step of a QM/DMD simulation involves 10,000 DMD steps followed by a loosely converged DFT geometry optimization. We have used QM/DMD to successfully study many aspects of metalloprotein behavior, including the effect of mutagenesis on structure and function,[50,52,53] metal-dependent catalytic activity,[10,54–56] redox functionality,[50,57] and recently, metal affinity.[10,58] Because of the sampling efficiency and capability of dynamically changing the metal coordination sphere, QM/DMD is suitable for building a CMA technology.

The exact form of the necessary free energy terms is another major complication in

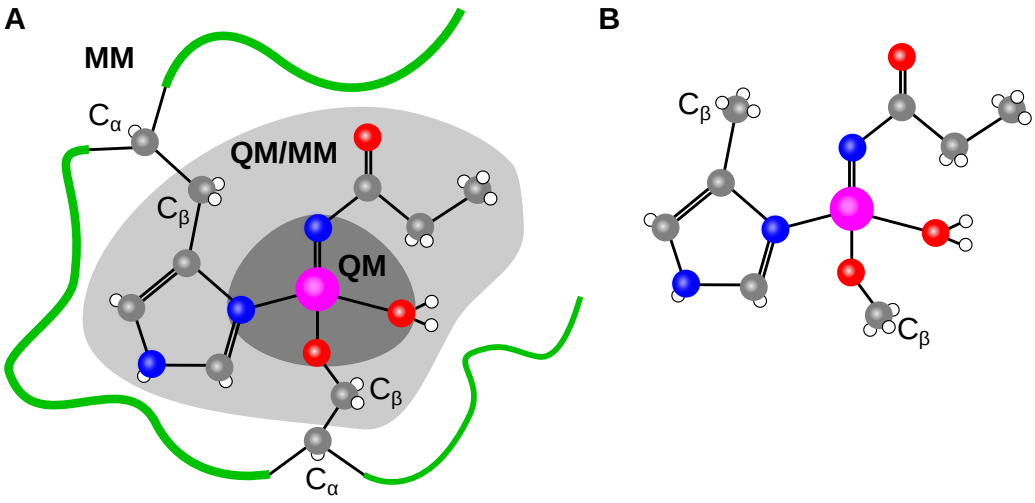

Figure 1: Diagrams demonstrating the active space of (A) QM/MM and (B) small cluster methods. QM/MM models the entire protein, with QM for the active site (the dark and light gray regions) and MM for the rest of the protein (the white region). In some forms of QM/MM, such as QM/DMD, there is an overlapping region treated with both QM and MM (in light gray) and MM modeling is only excluded from a small central region (in the case of this diagram, the metal and its first coordination sphere in dark gray). Small cluster methods, by contrast, only model the QM region.

CMA evaluation. One would think that metal affinities could be calculated from some combination of the free energies of the bound metalloprotein, and the apo-protein and the solvated metal ion. This is the approach of standard tools for free energy calculation in QM/MM biomolecular simulations[59,60] including adaptations of thermodynamic integration (TI)[61] and free energy perturbation (FEP).[62] These methods cannot be simply applied to metal binding processes. First, while it would be attractive for metal swapping, there is no accurate way to perform an alchemical transformation directly from one metal to another owing to their distinct electronic structures. Barring this, to obtain metal affinities these methods would need to model the process of metal binding from solution to protein. However, the accuracy of the free energies for these states will depend on the precision of evaluating the entropy change upon binding, which requires complete sampling of the conformational space of the protein both with and without the metal. Such full equilibration is practically impossible.[63] Additionally, evaluations of the free energy of the solvated metal ion requires expensive and laborious quantum mechanical treatment, explicit solvent, and

sufficient sampling of solvent configurations (on the order of 10^6). As metals are charged, ionic species, obtaining accurate, equilibrated results is more difficult than for the organic molecules that TI and FEP are applied to. Furthermore, this charged nature means that metal ion free energies cannot be directly obtained by experiment either.[64–66] In what follows we describe our CMA method that avoids all complications described in this paragraph. We will discuss several diverse applications of the method, its current limitations, and propose further directions to improve upon it. To the best of our knowledge, this technique is unprecedented.

2 Thermodynamic CMA Method

Our method calculates the relative metal binding free energy, $\Delta\Delta G$, with respect to one metal chosen as a reference. For many applications, relative free energies are sufficient as at least one metal is already known to bind. The approach combines QM/DMD sampling with a semi-empirical thermodynamic cycle that avoids ill-defined terms. First, we employ QM/DMD simulations run to convergence (on the order of 20-100 steps per replicate which is approximately 10-50 ns of sampling within DMD) of the protein with each considered metal. In the second step we determine the lowest energy QM region for each metal with optimization of the low-lying structures to tighter convergence and calculate its Gibbs free energy using the harmonic approximation. This approach concentrates on swiftly calculating a limited, but accurate free energy term for the region about the metal rather than pursuing an arduous and insufficiently accurate full protein free energy. Finally, we use these free energies in a thermodynamic cycle shown in Figure 2. The cycle consists of the metal ions going into the protein from a complex with a chelating agent (typically EDTA, which we exclusively used in all systems described in this article) rather than directly from solution. Hence, instead of using a dubious, calculated value for the free energy of a metal in solution, this cycle uses computationally tractable metal-chelator complexes. The free energies of

metal complexation from solution to the chelator are readily available from experiment. The final step of the cycle cancels the chelator terms through the computed free energies of metal exchange in the protein (from QM/DMD) and in the chelator complex (from *ab initio* or DFT calculations and harmonic vibrational entropies). Closing the thermodynamic cycle yields the $\Delta\Delta G$ of one metal, M_a, binding to the protein relative to the other metal, M_b. This means that when comparing the results of this method to experiment, only the trend can be reproduced, not the absolute free energies of metal binding.

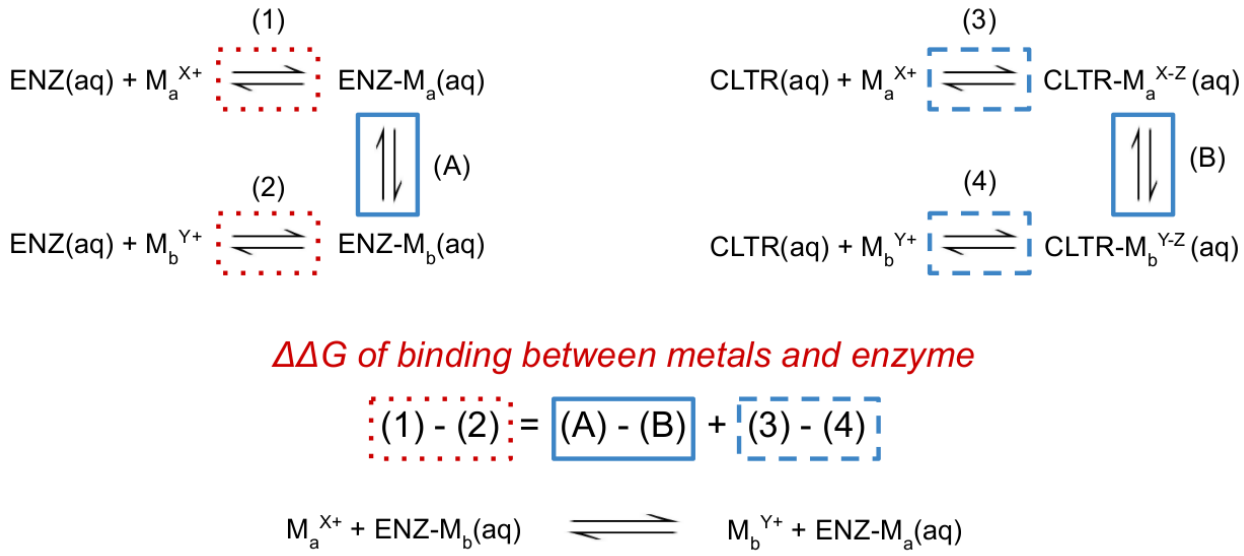

Figure 2: Thermodynamic cycles for the relative free energy of metal binding method. The left cycle of direct enzyme (ENZ) binding is intractable as the structure of free metal ions in solution is not defined (dashed red boxes). The right cycle uses experimentally available data for chelator (CLTR) binding to avoid this problem (dashed blue boxes). The sum of this cycle and the easily calculated transition from CLTR to the protein (solid blue boxes) gives the free energy of exchanging metals in the protein by canceling all the CLTR terms.

3 Method Benchmark

We have successfully applied the described CMA method to a series of problems of catalytic and biological relevance. To illustrate the method's performance and accuracy, we now describe several diverse examples, each with principally different biological functionality and

chemistry. We consider a mononuclear oxidase, a mononuclear metal-dependent hydrolase, and a metal transporter protein.

Acireductone dioxygenase (ARD) can tightly bind different metals and performs different reactions depending on which metal binds. The protein is involved in the methionine salvage pathway and acts on the substrate 1,2-dihydroxy-3-keto-5-(methylthio)pentene, oxidizing it to two possible sets of products.[67,68] ARD bound with Ni(II) catalyzes the formation of methylthiopropinate, while ARD bound with Fe(II) catalyzes the formation of 2-keto-4-methylthiobutyric acid, a precursor of methionine (Figure 3).[69] The bound metal does not change the structure of the protein, or the way in which the substrate binds to it, as we showed with QM/DMD. This means that the properties of the metal itself dictate catalytic selectivity. As such, ARD is the subject of many mechanistic studies.[55,70,71] We showed that the mechanistic bifurcation relies on the differences in charge transfer from the metal ligands, through the metal, and to the dioxygen bound to the substrate. Experimental binding studies show that ARD has an appreciable affinity for both Ni(II) and Fe(II).[72,73] The measured activity and metal binding affinities together demonstrate that both ARD reactive pathways are meaningful. The ARDs preference for the metal should then be context-dependent. Hence, the relative affinity of ARD to Fe(II) versus Ni(II) in the absence of other environmental factors is of interest.

The application of our CMA method to the catalytic metals in ARD, including Co(II), is illustrated in Table 1. To calculate the binding affinities of Fe(II), Ni(II), and Co(II) to ARD, we started with QM/DMD trajectories from our previous studies.[55] We selected the three lowest energy structures of the QM regions for each metal variant of ARD. We tested all feasible spin states of the metals with further geometry optimizations on these systems, looking for the multiplicity that minimizes the electronic energy. Our calculations showed that the multiplicity of Fe(II) was a singlet or quintet (depending on the structure), Ni(II) was a triplet, and Co(II) was a doublet. For each multiplicity we then performed frequency calculations and selected the lowest free energy among them. The calculations

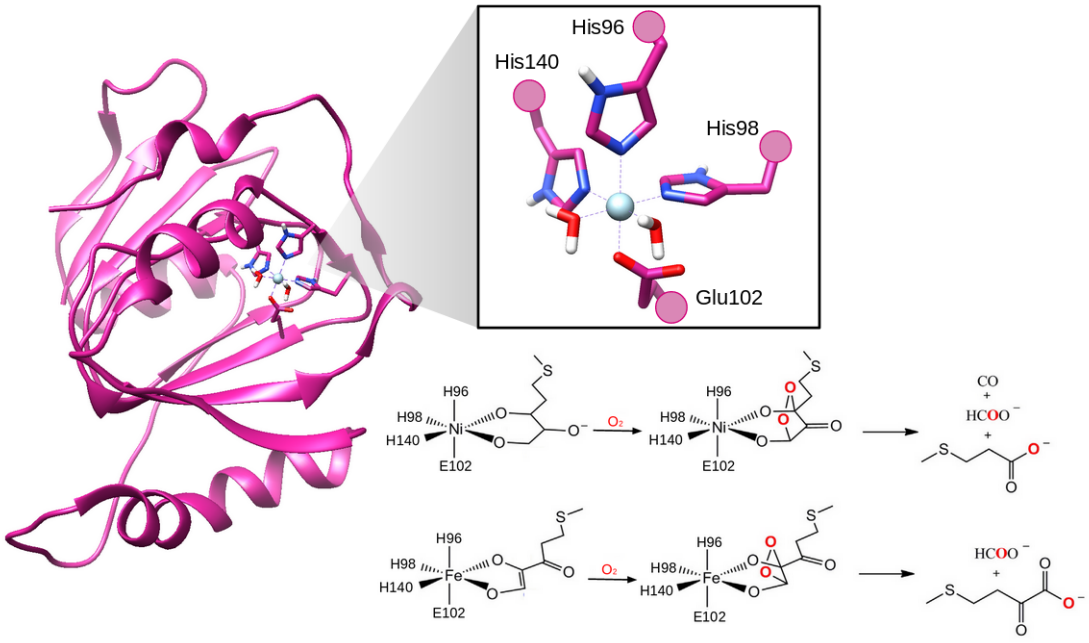

Figure 3: The structure of ARD (PDB ID: 1ZRR) and its active site, and the mechanisms of the metal-dependent reactions the protein can perform. The Ni(II) and Fe(II) bound forms of ARD preferentially bind different substrates and therefore perform different reactions.

were done with Turbomole (version 6.6).[74] The pure meta-GGA TPSS DFT functional[75] with the D3 dispersion correction[76] was used. The metal was treated with the triple-zeta basis set def2-TZVPP while all other atoms were treated with the double-zeta def2-SVP basis set. The conductor-like Screen Model (COSMO) with a constant dielectric of 20 was used to approximate the screening and solvation effects in the partially buried active site of the protein.[77] We selected this value based on the precedent of our previous, successful simulations of partially exposed active sites (such as the other examples we cover in this article). These settings are consistent with the initial QM/DMD runs. The results correctly capture that the affinity of the protein for Ni(II) is stronger than for Fe(II) and that ARD's affinity for Co(II) is about the same as for Fe(II). The quantitative difference between the computational and experimental values is about 1 to 2 kcal/mol (Table 1). Note that this approaches chemical accuracy (generally accepted as 1 kcal/mol) which is rarely achievable with DFT.[78,79] Given the many approximations needed along the way and despite the cancellation of errors in the relative calculations, the qualitative agreement with experiment we

obtained is still satisfying.

Table 1: Table of the experimental (Dai and Chai) and calculated binding affinities to ARD. The energies are relative to Fe(II), which correspondingly has a value of 0 kcal/mol. The experimental values here are based on Boltzmann weighted ratios of molar metal content.

	Fe(II)	Ni(II)	Co(II)
Dai (kcal/mol)	0.0	-1.23	-0.65
Chai (kcal/mol)	0.0	-0.28	N/A
Calc. (kcal/mol)	0.0	-3.76	0.38

Our next system is a histone deacetylase (HDAC), which is part of a class of enzymes that remove acetyl groups from histone lysines and potentially some nonhistone proteins.[80,81] Alongside histone acetyltransferases, which add acetyl groups, HDACs regulate how tightly histones bind to DNA and therefore gene regulation.[82–84] Overexpression of HDACs is associated with many pathologies, particularly cancer, while inhibition leads to the activation of genes related to growth arrest and tumor cells.[84,85] Consequently, many anti-cancer drugs are HDAC inhibitors.[86,87] Many of these bind to the transition metal center of their HDAC targets, including FDA approved suberanilohydroxamic acid (Vorinostat)[88] and FK228 (Romidepsin).[89] To reliably develop tighter binding drugs with computational methods, knowledge of which metal or metals bind to HDAC is necessary.

The catalytically relevant metals for histone deacetylases are not well understood. Historically, researchers assumed that HDACs are Zn(II) enzymes on the basis of X-ray structures and kinetic studies.[90,91] While Zn(II) is clearly a catalytically active metal in HDACs, as discussed earlier in this article, the promiscuity of metalloproteins means that crystallographic data does not preclude the relevance of other metals. Indeed, kinetic studies report significant activity for both Co(II) and Fe(II) in HDAC8, with Co(II) showing much higher activity than Zn(II).[9] This variety in metals that HDAC8 can use has important implications in traditional mechanistic studies.

Binding affinities from our method proved necessary to properly identify the catalytically relevant metals besides Zn(II) in HDAC8 and calculate their activities. Our group recently investigated the mechanism of HDAC8 and how it varies with physiologically abundant met-

als (Zn(II), Fe(II), Co(II), Mn(II), Ni(II), and Mg(II)) (Figure 4).[10] Pairing a traditional transition state search with QM/DMD simulations, we mapped the mechanism and calculated the activation barrier of the reaction for each metal. However, these results do not capture the experimental catalytic order and suggest that experimentally inactive Mn(II), Ni(II), and Mg(II) are reactive. We theorized that the binding affinities of these metals to HDAC8 contributes to their *in-vitro* catalytic activity. We calculated the $\Delta\Delta G$ for each metal and combined this with our computed barriers ($\Delta G^{\ddagger}$) to get a series of K_{rel}:

$$K_{rel} = exp\left(-\frac{\Delta G^{\ddagger}}{RT}\right) exp\left(-\frac{\Delta\Delta G_{binding}}{RT}\right)$$

which in contrast to the barriers, match the experimental catalytic order and identify Mn(II) and Mg(II) as inactive (Table 2). The K_{rel} of Ni(II) is the one outlier, with calculations suggesting that it is highly reactive, driven by its predicted high $\Delta\Delta G_{binding}$ rather than $\Delta G^{\ddagger}$. Ultimately, our study of HDAC8 demonstrates the utility of our metal binding $\Delta\Delta G$ method when the catalytic metal or metals of a natural metalloenzyme are not known.

Table 2: Table of the experimental k_{cat} and calculated K_{rel} values for HDAC8. While the exact values are not comparable, the qualitative order of the two catalytic measures match. Notice that Ni(II) is an exception, with the highest K_{rel} despite its experimental inactivity. Also note that Mg(II) and Mn(II) have K_{rel} that are many orders of magnitude lower than Co(II) meaning that they are consistent with their inactive experimental result.

	Co(II)	Zn(II)	Fe(II)	Ni(II)	Mn(II)	Mg(II)
Exp. k_{cat} (s^{-1})	1.2	0.90	0.48	N/A	N/A	N/A
Calc. K_{rel}	7.64×10^{-11}	1.27×10^{-11}	1.75×10^{-13}	1.89×10^{-8}	1.37×10^{-17}	1.46×10^{-23}

As an aside, we further hypothesize that in some cases the metal binding affinity could be a descriptor of enzymatic catalytic activity. Specifically, by the BEP principle,[92,93] the binding of the rate-determining intermediate to the active site should be neither too strong not too weak for the maximal catalytic activity to emerge. On the other hand, the stability of the active site itself and the metal ion in it should impact the stability of the intermediate of interest. That is because both the binding energy of the metal to its ligands, and the binding energy of the metal to the reaction intermediate depend on the energy and spatial extent of

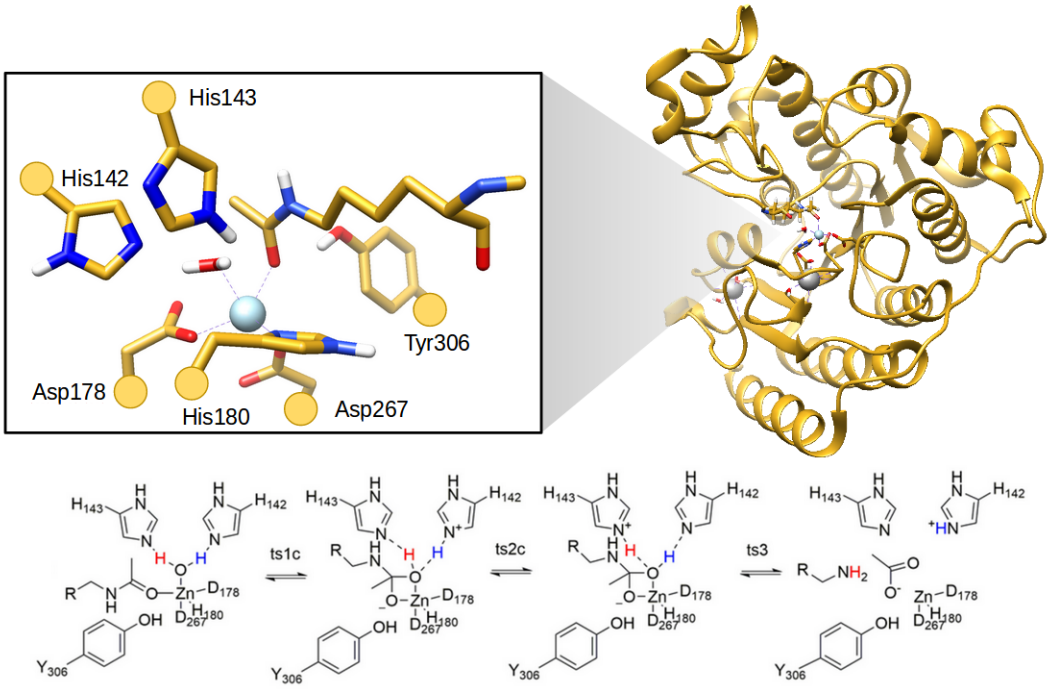

Figure 4: The structure of HDAC8 (PDB ID: 2V5W) and its active site with example substrate, and the most plausible mechanism of the deacetylation reaction it performs.

the orbitals of the metal. Therefore, there should be some relationship between the affinity of the protein to the metal and the catalytic activity of the metalloenzyme. We tested this conjecture using the computational data that we generated for the different metal variants of HDAC8, focusing just on the rate-determining, second step of the reaction (as shown in Figure 4). We excluded Mg(II) from the dataset, since it is known from the experiment to not bind appreciably to HDAC8. We correlate the $\Delta\Delta G$ of the metal ion binding to the protein to the Boltzmann weighted reaction barriers $e^{-E_a/RT}$ (which are the calculated k_{cat} normalized to remove the pre-exponential factor which we may assume to be approximately the same for all considered metals). The result is shown in Figure 5. We observe a classic volcano plot (a standard of heterogeneous catalysis analysis for the last 50 years)[94] that all metals obey, even Ni(II), demonstrating peak activity for a binding affinity around that of Zn(II). While we cannot assume that all metalloenzymes obey this sort of scaling relation, this demonstrates the utility of CMAs for yet another catalytic application.

Human serum transferrin (hTF) is an example of how CMAs could be used in a different

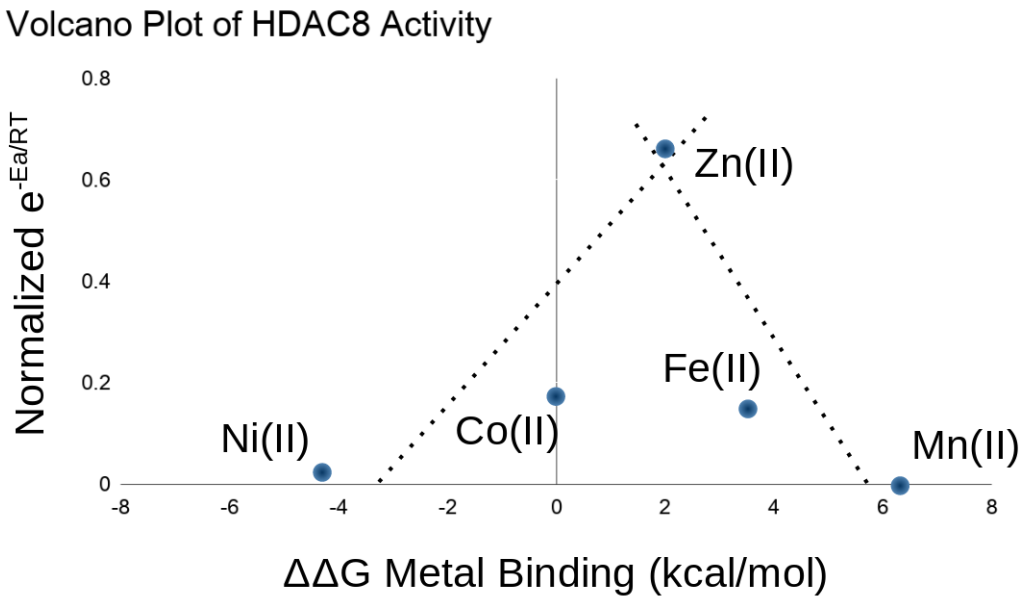

Figure 5: Volcano plot showing scaling relation of HDAC8 between binding $\Delta\Delta G$ and reaction rate. We calculated the reaction rates as the Boltzmann weighted ratios between each calculated k_{cat} and the Co(II) reference. The plotted values are normalized to remove the pre-exponential factor which we may assume to be approximately the same for all considered metals. Notice how even Ni(II) is consistent with this trend.

context. This protein is not catalytic, but is interesting for the purpose of this article because it can uptake and also release metals through pH dependent protein conformations with potentially profound implications in metal toxicology. The protein natively moves iron into cells by receptor-mediated endocytosis. Since it can cross the blood-brain barrier and its receptor is overexpressed in some cancer cells, hTF brings its cargo into particularly sensitive parts of the body.[95,96] Alarmingly, *in-vitro* binding studies show that hTF can bind other metals besides Fe(III),[97–99] including the potentially cytotoxic Ti(IV), Al(III), and Ga(III).[29,33] The promiscuity of hTF is of medical concern as these toxic metals are increasingly bioavailable with their use in modern industries, including in therapeutic drugs.[32,33,100,101] Previous studies provide some structural details on how hTF transports metals, but none access its full *in-vivo* activity. Two domains comprise the protein, each of which binds a single metal atom between two, highly similar subdomain lobes. Crystal structures and X-ray absorption fine structure spectroscopy studies of the N-terminal domain suggest that the lobes hinge open

in the endosome environment (Figure 6).[102] Such a conformational change encourages metal release. Previous classical MD simulations could access protein opening,[103,104] but not the thermodynamic data about the metal release.

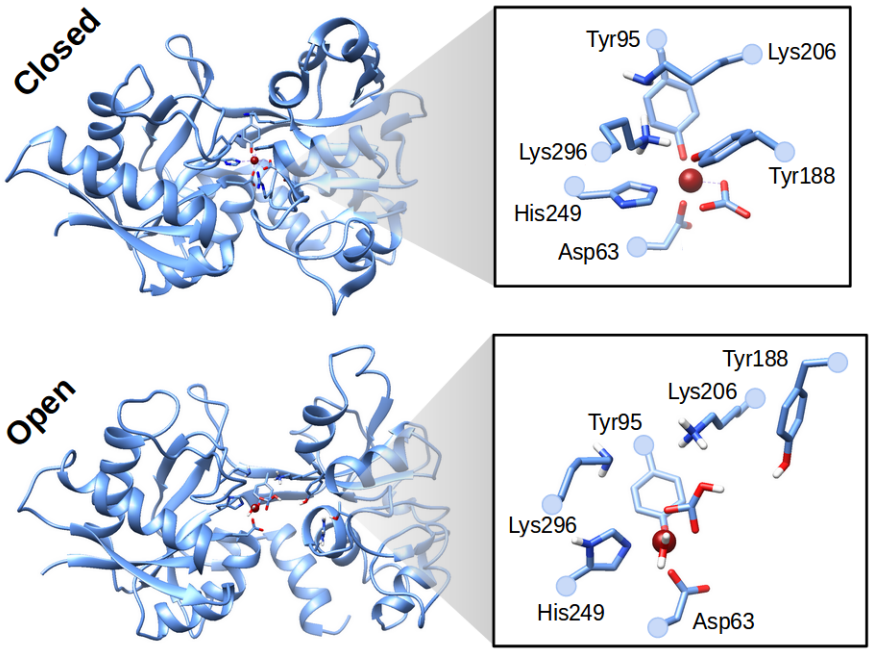

Figure 6: The structure of hTF and its active site in the closed and open forms of the protein. The closed form is associated with the pH of the blood serum, while the open form is associated with the low pH conditions of the endosome. The closed structure was obtained by X-ray crystallography (PDB ID: 3V83), while the open structure was obtained from computational studies.[58]

We used our CMA method to get the first insight into the toxic metal transport abilities of hTF *in-vivo* conformational states. We calculated metal binding affinities relative to physiological Fe(III) for Ti(IV), Co(III), Ga(III), Cr(III), Fe(II), Zn(II) in both uptake and release implicated forms of hTF (Figure 7).[58] The order of the binding affinities in the uptake form of the protein are qualitatively consistent with experiment. Accordingly, as Ti(IV), Co(III), Ga(III), and Cr(III) demonstrate $\Delta\Delta G$ that are negative or about 0 in this form, our results show that hTF can uptake these given metals competitively with Fe(III). In contrast, the results for the release states vary for these metals, with Co(III) and Cr(III) reporting consistently large $\Delta\Delta G$, but Ti(IV) and Ga(III) reporting small or negative $\Delta\Delta G$

in one form. This suggests that hTF releases Co(III) and Cr(III) much more readily than Fe(III), but releases Ti(IV) and Ga(III) about as readily as Fe(III). Since Ti(IV) and Ga(III) strongly bind to both the closed and open states, these cytotoxic metals may sequester some of the protein. Further, our study identified the protein residues that are most likely to be responsible for opening and closing at changing pH, as well as a collection of geometric and electronic factors that are responsible for the different affinities of hTF to the studied metals.

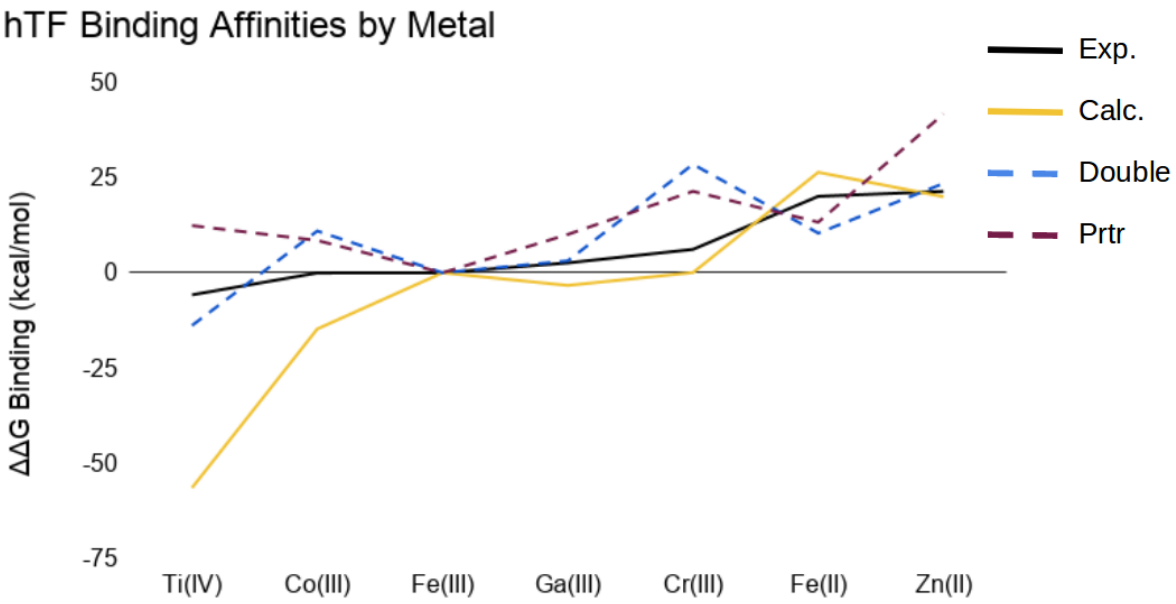

Figure 7: Metal binding affinities to hTF from experiment (solid black line) and from our method. All affinities are relative to Fe(III), which correspondingly has a value of 0 kcal/mol for all lines. The experimental values are based on Boltzmann weighted ratios of binding constants. The solid yellow line are the values from uptake form of the protein. Notice that it matches the shape and order of the experimental line. Furthermore, note that Ti(IV), Co(III), Ga(III), and Cr(III) all have values that are negative or around 0 for this line. The dashed light blue and dark purple lines are the values from the release forms of the protein (called 'Double' and 'Prtr'). The difference between the two forms is minor, but significant; the structure represented with the light blue line contains an additional water molecule in its active site. Notice that for at least one of the dashed lines, both Ti(IV) and Ga(III) bind about the same or better than Fe(III).

4 Limitations and Outlook

Further research into CMA methods is important, especially as our method is not without limitations. Its reliance on chelating agents introduces other problems besides limiting calculations to referential $\Delta\Delta G$. The best way to calculate the thermodynamic terms involving the chelating agent is unclear. Experimental stability constants for EDTA and many related chelating agents are fortunately available for most metals in their common oxidation states.[105] Unfortunately, the corresponding structures of these metal complexes are not fully known, and they are necessary to accurately calculate the free energy associated with the transition from the chelator complex to the protein. In the studies we discuss above, we assume full chelation of each metal with no other ligands in the complexes. This makes most metals conform to an octahedral geometry. This is likely fine for large transition metals, but breaks down for small and low charge metals such as Li(I) and Mg(II). Indeed, crystallographic studies of Mg-EDTA binding show that a water molecule is also a ligand in the complex.[106] One way to mitigate these problems would be a benchmark study of a wide range of chelators on a system that has been experimentally well characterized for many metals. Calculating the set of $\Delta\Delta G$ for each chelator without varying any other parameters would reveal which chelator can be used most accurately for each metal.

Our method is also limited to proteins which undergo only minor conformational changes upon the binding of different metals. The first concern here is that the QM regions must share the same atoms besides the metal center to satisfy the thermodynamic cycle. Metals that bind entirely different sites on a protein are consequentially inaccessible to our current method. A second concern largely involves computational scaling, as significant rearrangement (like refolding) upon metal binding requires even more expensive structural sampling in order to accurately assess the entropy component of $\Delta\Delta G$. While this is a general problem with protein and metalloprotein simulations, enhanced sampling for the specific purpose of metal binding affinities would be impactful. Solutions to both of these concerns would render many systems more accessible, particularly metal chaperones as these proteins can

adopt different folds for different metals.[24]

Our current CMA is also limited in accuracy and chemical scope by its use of DFT. Traditionally, DFT struggles with multireference systems, where one Slater determinant or configuration state function is insufficient, especially metal clusters. Certain post-Hartree-Fock wavefunction methods can appropriately treat these cases and are particularly important for accurate energies. There is already much discussion on the use of these tools in heterogeneous catalysis.[107] As multireference post-Hartree-Fock methods tend to be computationally intensive, multiconfiguration pair-density functional theory that blends wavefunction methods with DFT is promising for CMA applications because of their affordability.[108] Future CMAs could use such methods specifically for the free energy calculations on the QM region to obtain more accurate energies without increasing computational cost too drastically.

Further advancements in CMA methods would greatly propel understanding of natural metalloenzymes and the design of new ArMs. Such techniques could determine the catalytically relevant metals in natural metalloenzymes, which cannot be taken for granted from crystal structures. CMA calculations would be indispensable in the effort to better understand metal transport pathways throughout the body, especially with regards to metal toxicology. In the design of ArMs, replacing the bound metal in an existing metalloprotein scaffold can introduce new functions, often inaccessible to current design methodologies like directed evolution. Placing a metal into a specifically designed artificial scaffold is also an attractive opportunity for ArMs catalysis. For all such design tasks, it is critical to assess the metal affinity and its ability to outperform other metals that might be present in the synthesis conditions. New tools such as CMAs will expand the catalytic space of metalloenzymes.

Acknowledgement

We thank the Institute for Digital Research and Education at UCLA and the Extreme Science and Engineering Discovery Environment for supercomputer time. The support for this work over the years came from the NSF-CAREER Award CHE-1351968, NSFCHE-1903808, and NIGMS 1R01GM134047.

References

(1) Levy, R.; Edelman, M.; Sobolev, V. *Proteins: Structure, Function, and Bioinformatics* **2009**, *76*, 365–374.

(2) Passerini, A.; Lippi, M.; Frasconi, P. *Nucleic Acids Research* **2011**, *39*, W288–W292.

(3) Sodhi, J. S.; Bryson, K.; McGuffin, L. J.; Ward, J. J.; Wernisch, L.; Jones, D. T. *Journal of Molecular Biology* **2004**, *342*, 307–320.

(4) Zheng, H.; Chordia, M. D.; Cooper, D. R.; Chruszcz, M.; Müller, P.; Sheldrick, G. M.; Minor, W. *Nature Protocols* **2014**, *9*, 156.

(5) Giambasu, G. M.; Case, D. A.; York, D. M. *Journal of the American Chemical Society* **2019**, *141*, 2435–2445.

(6) Kepp, K. P. *Coordination Chemistry Reviews* **2017**, *344*, 363–374.

(7) Valdez, C. E.; Smith, Q. A.; Nechay, M. R.; Alexandrova, A. N. *Accounts of Chemical Research* **2014**, *47*, 3110–3117.

(8) Tripp, B. C.; Bell, C. B.; Cruz, F.; Krebs, C.; Ferry, J. G. *Journal of Biological Chemistry* **2004**, *279*, 6683–6687.

(9) Gantt, S. L.; Gattis, S. G.; Fierke, C. A. *Biochemistry* **2006**, *45*, 6170–6178.

(10) Nechay, M. R.; Gallup, N. M.; Morgenstern, A.; Smith, Q. A.; Eberhart, M. E.; Alexandrova, A. N. *The Journal of Physical Chemistry B* **2016**, *120*, 5884–5895.

(11) Zhu, J.; Dizin, E.; Hu, X.; Wavreille, A.-S.; Park, J.; Pei, D. *Biochemistry* **2003**, *42*, 4717–4726.

(12) Rajagopalan, P. R.; Yu, X. C.; Pei, D. *Journal of the American Chemical Society* **1997**, *119*, 12418–12419.

(13) Renata, H.; Wang, Z. J.; Arnold, F. H. *Angewandte Chemie International Edition* **2015**, *54*, 3351–3367.

(14) Hyster, T. K.; Ward, T. R. *Angewandte Chemie International Edition* **2016**, *55*, 7344–7357.

(15) Fasan, R.; Meharenna, Y. T.; Snow, C. D.; Poulos, T. L.; Arnold, F. H. *Journal of Molecular Biology* **2008**, *383*, 1069–1080.

(16) Lewis, J. C.; Bastian, S.; Bennett, C. S.; Fu, Y.; Mitsuda, Y.; Chen, M. M.; Greenberg, W. A.; Wong, C.-H.; Arnold, F. H. *Proceedings of the National Academy of Sciences* **2009**, *106*, 16550–16555.

(17) Rentmeister, A.; Brown, T. R.; Snow, C. D.; Carbone, M. N.; Arnold, F. H. *ChemCatChem* **2011**, *3*, 1065–1071.

(18) Yu, F.; Cangelosi, V. M.; Zastrow, M. L.; Tegoni, M.; Plegaria, J. S.; Tebo, A. G.; Mocny, C. S.; Ruckthong, L.; Qayyum, H.; Pecoraro, V. L. *Chemical Reviews* **2014**, *114*, 3495–3578.

(19) Reetz, M. T. *Accounts of Chemical Research* **2019**, *52*, 336–344.

(20) Prier, C. K.; Arnold, F. H. *Journal of the American Chemical Society* **2015**, *137*, 13992–14006.

(21) Natoli, S. N.; Hartwig, J. F. *Accounts of Chemical Research* **2019**, *52*, 326–335.

(22) Finney, L. A.; O'Halloran, T. V. *Science* **2003**, *300*, 931–936.

(23) Tottey, S.; Harvie, D. R.; Robinson, N. J. *Accounts of Chemical Research* **2005**, *38*, 775–783.

(24) Tottey, S.; Waldron, K. J.; Firbank, S. J.; Reale, B.; Bessant, C.; Sato, K.; Cheek, T. R.; Gray, J.; Banfield, M. J.; Dennison, C.; Robinson, N. J. *Nature* **2008**, *455*, 1138.

(25) Foster, A. W.; Osman, D.; Robinson, N. J. *Journal of Biological Chemistry* **2014**, *289*, 28095–28103.

(26) Xiao, Z.; Wedd, A. G. *Natural Product Reports* **2010**, *27*, 768–789.

(27) Pieczenik, S. R.; Neustadt, J. *Experimental and Molecular Pathology* **2007**, *83*, 84–92.

(28) Ibrahim, D.; Froberg, B.; Wolf, A.; Rusyniak, D. E. *Clinics in Laboratory Medicine* **2006**, *26*, 67–97.

(29) Exley, C.; Burgess, E.; Day, J. P.; Jeffery, E. H.; Yokel, R. A. *Journal of Toxicology and Environmental Health Part A* **1996**, *48*, 569–584.

(30) Tinoco, A. D.; Thomas, H. R.; Incarvito, C. D.; Saghatelian, A.; Valentine, A. M. *Proceedings of the National Academy of Sciences* **2012**, *109*, 5016–5021.

(31) Guo, M.; Sun, H.; McArdle, H. J.; Gambling, L.; Sadler, P. J. *Biochemistry* **2000**, *39*, 10023–10033.

(32) Jakupec, M. A.; Keppler, B. K. *Current Topics in Medicinal Chemistry* **2004**, *4*, 1575–1583.

(33) Exley, C. *Environmental Science: Processes & Impacts* **2013**, *15*, 1807–1816.

(34) Gaggelli, E.; Kozlowski, H.; Valensin, D.; Valensin, G. *Chemical Reviews* **2006**, *106*, 1995–2044.

(35) Hoops, S. C.; Anderson, K. W.; Merz Jr, K. M. *Journal of the American Chemical Society* **1991**, *113*, 8262–8270.

(36) Dal Peraro, M.; Spiegel, K.; Lamoureux, G.; De Vivo, M.; DeGrado, W. F.; Klein, M. L. *Journal of Structural Biology* **2007**, *157*, 444–453.

(37) Neves, R. P.; Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Journal of Chemical Theory and Computation* **2013**, *9*, 2718–2732.

(38) Cho, A. E.; Goddard III, W. A. *Metalloproteins: theory, calculations, and experiments*; CRC Press: Boca Raton, FL, 2015.

(39) Dal Peraro, M.; Vila, A. J.; Carloni, P.; Klein, M. L. *Journal of the American Chemical Society* **2007**, *129*, 2808–2816.

(40) Zhang, J.; Yang, W.; Piquemal, J.-P.; Ren, P. *Journal of Chemical Theory and Computation* **2012**, *8*, 1314–1324.

(41) Panteva, M. T.; Giambasu, G. M.; York, D. M. *The Journal of Physical Chemistry B* **2015**, *119*, 15460–15470.

(42) Rydberg, P.; Sigfridsson, E.; Ryde, U. *JBIC Journal of Biological Inorganic Chemistry* **2004**, *9*, 203–223.

(43) Tantillo, D. J. *Organic Letters* **2010**, *12*, 1164–1167.

(44) Kries, H.; Blomberg, R.; Hilvert, D. *Current Opinion in Chemical Biology* **2013**, *17*, 221–228.

(45) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. *Angewandte Chemie International Edition* **2013**, *52*, 5700–5725.

(46) Vaissier Welborn, V.; Head-Gordon, T. *Chemical Reviews* **2018**, *119*, 6613–6630.

(47) Blomberg, M. R.; Borowski, T.; Himo, F.; Liao, R.-Z.; Siegbahn, P. E. *Chemical Reviews* **2014**, *114*, 3601–3658.

(48) Ryde, U. *Methods in Enzymology*; Elsevier, 2016; Vol. 577; pp 119–158.

(49) Ahmadi, S.; Barrios Herrera, L.; Chehelamirani, M.; Hostaš, J.; Jalife, S.; Salahub, D. R. *International Journal of Quantum Chemistry* **2018**, *118*, e25558.

(50) Sparta, M.; Shirvanyants, D.; Ding, F.; Dokholyan, N. V.; Alexandrova, A. N. *Biophysical Journal* **2012**, *103*, 767–776.

(51) Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. *Structure* **2008**, *16*, 1010–1018.

(52) Valdez, C. E.; Morgenstern, A.; Eberhart, M. E.; Alexandrova, A. N. *Physical Chemistry Chemical Physics* **2016**, *18*, 31744–31756.

(53) Reilley, D. J.; Popov, K.; Dokholyan, N. V.; Alexandrova, A. N. *The Journal of Physical Chemistry B* **2019**, *123*, 4534–4539.

(54) Valdez, C. E.; Alexandrova, A. N. *The Journal of Physical Chemistry B* **2012**, *116*, 10649–10656.

(55) Sparta, M.; Valdez, C. E.; Alexandrova, A. N. *Journal of Molecular Biology* **2013**, *425*, 3007–3018.

(56) Valdez, C. E.; Gallup, N. M.; Alexandrova, A. N. *Chemical Physics Letters* **2014**, *604*, 77–82.

(57) Nedd, S.; Redler, R. L.; Proctor, E. A.; Dokholyan, N. V.; Alexandrova, A. N. *Journal of Molecular Biology* **2014**, *426*, 4112–4124.

(58) Reilley, D. J.; Fuller III, J. T.; Nechay, M. R.; Victor, M.; Li, W.; Ruberry, J. D.; Mujika, J. I.; Lopez, X.; Alexandrova, A. N. *Pending* **2019**,

(59) Kästner, J.; Senn, H. M.; Thiel, S.; Otte, N.; Thiel, W. *Journal of Chemical Theory and Computation* **2006**, *2*, 452–461.

(60) Giese, T. J.; York, D. M. *Journal of Chemical Theory and Computation* **2019**, *15*, 5543–5562.

(61) Kirkwood, J. G. *The Journal of Chemical Physics* **1935**, *3*, 300–313.

(62) Zwanzig, R. W. *The Journal of Chemical Physics* **1954**, *22*, 1420–1426.

(63) Genheden, S.; Ryde, U. *Physical Chemistry Chemical Physics* **2012**, *14*, 8662–8677.

(64) Guggenheim, E. *The Journal of Physical Chemistry* **1929**, *33*, 842–849.

(65) Klotz, I. M.; Rosenberg, R. M. *Chemical Thermodynamics*; Wiley: New York, NY, 1994.

(66) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *The Journal of Physical Chemistry B* **2006**, *110*, 16066–16081.

(67) Myers, R. W.; Wray, J.; Fish, S.; Abeles, R. *Journal of Biological Chemistry* **1993**, *268*, 24785–24791.

(68) Oram, S. W.; Ai, J.; Pagani, G. M.; Hitchens, M. R.; Stern, J. A.; Eggener, S.; Pins, M.; Xiao, W.; Cai, X.; Haleem, R.; Jiang, F.; Pochapsky, T. C.; Hedstrom, L.; Wang, Z. *Neoplasia* **2007**, *9*, 643.

(69) Dai, Y.; Wensink, P. C.; Abeles, R. H. *Journal of Biological Chemistry* **1999**, *274*, 1193–1195.

(70) Wray, J. W.; Abeles, R. H. *Journal of Biological Chemistry* **1995**, *270*, 3147–3153.

(71) Borowski, T.; Bassan, A.; Siegbahn, P. E. *Journal of Molecular Structure: THEOCHEM* **2006**, *772*, 89–92.

(72) Dai, Y.; Pochapsky, T. C.; Abeles, R. H. *Biochemistry* **2001**, *40*, 6379–6387.

(73) Chai, S. C.; Ju, T.; Dang, M.; Goldsmith, R. B.; Maroney, M. J.; Pochapsky, T. C. *Biochemistry* **2008**, *47*, 2428–2438.

(74) Turbomole, V. 6.6; Turbomole Gmbh: Karlsruhe, Germany, 2014.

(75) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *The Journal of Chemical Physics* **2003**, *119*, 12129–12137.

(76) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *The Journal of Chemical Physics* **2010**, *132*, 154104.

(77) Klamt, A. *The Journal of Physical Chemistry* **1995**, *99*, 2224–2235.

(78) Cramer, C. J.; Truhlar, D. G. *Physical Chemistry Chemical Physics* **2009**, *11*, 10757–10816.

(79) Mardirossian, N.; Head-Gordon, M. *Molecular Physics* **2017**, *115*, 2315–2372.

(80) Kouzarides, T. *The EMBO Journal* **2000**, *19*, 1176–1179.

(81) Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. *Science* **2009**, *325*, 834–840.

(82) Phillips, D. *Biochemical Journal* **1963**, *87*, 258.

(83) Allfrey, V.; Faulkner, R.; Mirsky, A. *Proceedings of the National Academy of Sciences* **1964**, *51*, 786–794.

(84) Gallinari, P.; Di Marco, S.; Jones, P.; Pallaoro, M.; Steinkühler, C. *Cell Research* **2007**, *17*, 195.

(85) Haberland, M.; Montgomery, R. L.; Olson, E. N. *Nature Reviews Genetics* **2009**, *10*, 32.

(86) Marks, P. A.; Breslow, R. *Nature Biotechnology* **2007**, *25*, 84.

(87) West, A. C.; Johnstone, R. W. *The Journal of Clinical Investigation* **2014**, *124*, 30–39.

(88) Lobera, M. et al. *Nature Chemical Biology* **2013**, *9*, 319.

(89) Furumai, R.; Matsuyama, A.; Kobashi, N.; Lee, K.-H.; Nishiyama, M.; Nakajima, H.; Tanaka, A.; Komatsu, Y.; Nishino, N.; Yoshida, M.; Horinouchi, S. *Cancer Research* **2002**, *62*, 4916–4921.

(90) Finnin, M. S.; Donigian, J. R.; Cohen, A.; Richon, V. M.; Rifkind, R. A.; Marks, P. A.; Breslow, R.; Pavletich, N. P. *Nature* **1999**, *401*, 188.

(91) Drummond, D. C.; Noble, C. O.; Kirpotin, D. B.; Guo, Z.; Scott, G. K.; Benz, C. C. *Annual Review of Pharmacology and Toxicology* **2005**, *45*, 495–528.

(92) Bell, R. P. *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences* **1936**, *154*, 414–429.

(93) Evans, M.; Polanyi, M. *Transactions of the Faraday Society* **1936**, *32*, 1333–1360.

(94) Balandin, A. *Advances in Catalysis*; Elsevier, 1969; Vol. 19; pp 1–210.

(95) Li, H.; Qian, Z. M. *Medicinal Research Reviews* **2002**, *22*, 225–250.

(96) Gupta, Y.; Jain, A.; Jain, S. K. *Journal of Pharmacy and Pharmacology* **2007**, *59*, 935–940.

(97) Li, H.; Sadler, P. J.; Sun, H. *European Journal of Biochemistry* **1996**, *242*, 387–393.

(98) Tinoco, A. D.; Valentine, A. M. *Journal of the American Chemical Society* **2005**, *127*, 11218–11219.

(99) Tinoco, A. D.; Incarvito, C. D.; Valentine, A. M. *Journal of the American Chemical Society* **2007**, *129*, 3444–3454.

(100) Sun, H.; Li, H.; Sadler, P. J. *Chemical Reviews* **1999**, *99*, 2817–2842.

(101) Cini, M.; Bradshaw, T. D.; Woodward, S. *Chemical Society Reviews* **2017**, *46*, 1040–1051.

(102) Baker, H. M.; Nurizzo, D.; Mason, A. B.; Baker, E. N. *Acta Crystallographica Section D: Biological Crystallography* **2007**, *63*, 408–414.

(103) Mujika, J. I.; Escribano, B.; Akhmatskaya, E.; Ugalde, J. M.; Lopez, X. *Biochemistry* **2012**, *51*, 7017–27.

(104) Rinaldo, D.; Field, M. J. *Biophysical Journal* **2003**, *85*, 3485–3501.

(105) Dojindo Molecular Technologies, I. Metal Chelates. Accessed: 2019-09-18.

(106) Passer, E.; White, J.; Cheng, K. *Inorganica Chimica Acta* **1977**, *24*, 13–23.

(107) Gaggioli, C. A.; Stoneburner, S. J.; Cramer, C. J.; Gagliardi, L. *ACS Catalysis* **2019**, *9*, 8481–8502.

(108) Gagliardi, L.; Truhlar, D. G.; Li Manni, G.; Carlson, R. K.; Hoyer, C. E.; Bao, J. L. *Accounts of Chemical Research* **2016**, *50*, 66–73.

Graphical TOC Entry